Study on Data Compression and Reduction of the Aviation Network Based on Multi-resolution Wavelet Analysis

Yao Hong Guang

School of Air Transportation / Flying, Shanghai University of Engineering and Science, Shanghai, China

Email address:
yhg1yhg@sina.com

Abstract: This paper proposes complex network data compression idea based on the multi-resolution wavelet decomposition theory, analyzes the concrete form of wavelet basis choice and wavelet decomposition, and then puts forward the determination method of network decomposition levels and parameters reduction method after decomposition. The empirical study shows that four-level wavelet decomposition for Chinese aviation network adjacency matrix is carried out by Haar wavelet basis, and the lowest frequency sub-band is 10×10 matrix. Moreover, the average degree, the average shortest path length and clustering coefficient of original network are restored in the lowest frequency sub-band after decomposition.

Keywords: Aviation Network, Multi-resolution Wavelet Decomposition, Complex Network,
Data Compression and Reduction

1. Introduction

Aviation network can be regarded as the complex network composed of cities which the airports stand in and the air routes which link these cities, and navigable cities and air routes constitute the basic framework of aviation network. In the aviation network, one air route connects two navigable cities, which realizes the bidirectional air transport between the cities, so the network can be considered as undirected network. Aviation network has the characteristics of typical complex network [1]. It is the premise and foundation for the network analysis to measure the statistical characteristic parameters of aviation network. However, there appears considerable difficulty for the empirical study of aviation network characteristics due to the complexity and dynamics of aviation network node and air route. Wavelet multi-resolution decomposition is currently main technical means for the complex network. Marr. D proposed wavelet multi-resolution theory by simulating the human visual system for the first time [2]. Furthermore, the study from Fan J et al. showed that multi-resolution decomposition technology has been an effective tool for studying complex network structure properties.[3] The adjacency matrix $A(p) = a_{ij}$ should be firstly determined for the wavelet decomposition of a network. If there exists connection edge between i and j, then $a_{ij} = 1$, if not, $a_{ij} = 0$. The wavelet decomposition is carried out for

$A(p)$, and the appropriate wavelet basis is chosen. The high-pass filter H and low-pass filter L are used for filtering along the horizontal and vertical directions of network. And then four subsegments (HH, HL, LH, and LL) are output. The wavelet decomposition proceeds in the low frequency sub-band until the decomposition requirements are satisfied [4], as shown in Figure 1.
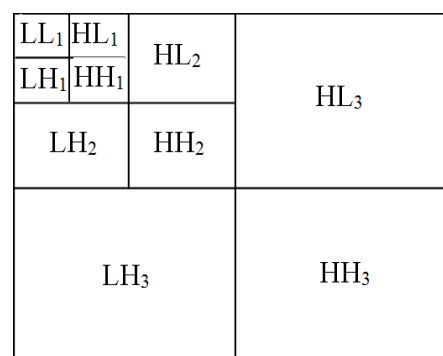

Fig. 1. Three-level wavelet decomposition diagram.

The study shows that multi-resolution wavelet decomposition has quite good energy tightness [5]. Most of the system energy concentrates on the low frequency sub-band, but the high frequency sub-band has little or no energy. The information contained by low frequency sub-band LL_1 can be regarded as the coarse graining description of network,

other sub-bands include information with higher resolution, and different sub-bands can describe the network with different accuracy. This analysis thinking is similar to visual processing of human eyes [6]. Multi-resolution wavelet decomposition provides a natural way to analyze complex network. Namely, firstly, the network structure is analyzed by the coarse resolution, and then the analytical precision is gradually increased according to the requirements [7]. Hence, this method has excellent adaptation for processing fairly complex system of aviation network.

2. Major Parameters Analysis of Chinese Aviation Network

Network parameters intensively reflect the basic situation of the network, which is the foundation of network analysis.

2.1. Adjacency Matrix and Node Degree of Aviation Network

(1) Adjacency matrix of aviation network

The establishment of adjacency matrix is the basic work of complex network study [8]. The adjacency matrix of the aviation network stores the mutual connection information of the network in the form of matrix. If there is conjoint air routes between the navigable cities i and j in the aviation network, the element of adjacency matrix $a_{ij} = 1$, or else $a_{ij} = 0$.

(2) Degree

In aviation network, the degree k_i of node city i is defined as the total number of air routes connecting with the node city i. Namely, the greater the degree of one city is, the more the routes connecting with the city are [9].

2.2. Average Shortest Path of Aviation Network

The shortest path is the number of air routes included in the shortest routes between two node cities in the aviation network. And average shortest path length is the number of average shortest routes of all the node cities pairs in aviation network, which reflects the needed average least transit times for the passengers and goods from one node to another one [10]. The calculation method for the average shortest path length L of the aviation network is as shown in equation (1).

$$L = \frac{\sum L_{ij}}{\frac{1}{2}N(N+1)} \qquad (1)$$

Where N is the number of nodes in the aviation network, L_{ij} is the number of air routes of the shortest path between the node cities i and j.

2.3. Clustering Coefficient

The clustering coefficient C_i of the node i can be used to describe the dense degree that the node i of aviation network connects with the other nodes; its value can be calculated by the equation (2).

$$C_i = \frac{\text{The number of triangles connecting with node i}}{\text{The number of triples connecting with node i}} \qquad (2)$$

The clustering coefficient C of aviation network is the average value of clustering coefficient Ci of all the nodes i. The greater value reflects that the distribution of the network routes is more uneven, namely, local area density is greater, but some regional routes are sparse [11].

3. Wavelet Decomposition of Chinese Aviation Network

3.1. Wavelet Decomposition of Aviation Network Based on the Haar Wavelet Conversion

(1) Applicability analysis of Haar wavelet

Haar function is a group of function set, from which the Haar wavelet is derived, which is the simplest binary wavelet function with discontinuous time-domain, and it is defined as follows.

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{other} \end{cases} \qquad (3)$$

Due to the characteristic of Haar wavelet function, it has a good applicability for the decomposition and conversion of binary system or 0-1 matrix. The adjacency matrix is used to store the data information for the aviation network. The adjacency matrix is a 0-1 matrix, which can reflect the connection condition between i and j of any two points in the network. Therefore, the Haar wavelet is the ideal wavelet for aviation network decomposition.

(2) Haar wavelet decomposition of aviation network

Multi-resolution decomposition is carried out for the aviation network by using the Haar wavelet and adjacency matrix. The first level decomposition is that the original connection matrix is decomposed into a low frequency part (relative to the average) and three high frequency parts (corresponding to the details in the horizontal, vertical and diagonal locations). The process mentioned above is repeated to realize aviation network of the multi-layer wavelet decomposition.

It is assumed that the adjacent matrix of aviation network is A, and the r-level wavelet decomposition of A is denoted by $\tilde{A^r}$. N nodes in aviation network are divided into 2^r groups, and the ith group includes $N/2^r$ nodes in sequence[12], namely

$$2^r(r-1)+1, \ 2^r(r-1)+2, \ 2^r(r-1)+3, \ ; \ i=1, \ 2, \ .., \ N/2^r \qquad (4)$$

The element value of adjacent matrix by r-level wavelet transform can be derived by the formula (4).

$$\tilde{a}_{ij}^r = \frac{1}{2^r}(\tilde{a}_{2i-1,2j-1}^r + \tilde{a}_{2i-1,2j}^r + \tilde{a}_{2i,2j-1}^r + \tilde{a}_{2i,2j}^r) = \frac{1}{2^r}\sum_{s=2^r(i-1)+1}^{2^r i}\sum_{t=2^r(j-1)+1}^{2^r j} a_{st} \quad i,\ j=1,\ 2,\ ,N/2^r \tag{5}$$

Hence, r-level wavelet transform of aviation network adjacent matrix A corresponds to the divided 2^r groups of network nodes, and each group includes $N/2^r$ nodes in sequence. The diagonal element of the lowest frequency sub-band LL_1 of wavelet transform signifies the normalized number of internal edges in each group, while the LL_1 non-diagonal element expresses the normalized number of edges connecting different groups, and the normalized constant is 2^r. This research shows that the elements in the lowest frequency sub-band LL_1 matrix after network decomposition contains most of the original network information [13]. Hence, the original adjacency matrix can be replaced by the lowest frequency sub-band LL_1 to simplify network.

3.2. Parameter Analysis after the Network Decomposition

3.2.1. Connection Property Analysis

The off-diagonal element $\tilde{a}_{ij}^r$ of the network low frequency sub-band LL_1 after wavelet decomposition depicts the connection property between the ith group and the jth group in the primary network. $\tilde{a}_{ij}^r \neq 0$ means that if and only if there is one edge at least between the two groups, and the greater value of $\tilde{a}_{ij}^r$ indicates that the number of edges is more between the two groups. In a similar way, the diagonal element $\tilde{a}_{ii}^r$ of LL_1 describes the connection property in the interior of the ith group. The greater value of $\tilde{a}_{ii}^r$ is, the more the number of the interior edges is[14].

3.2.2. Average Shortest Path Length Analysis

One node of new network after wavelet corresponds to one group of the original network. One edge between two different nodes i and j in new network is equivalent to one or more edges between the ith group and jth group in the original network. If the average shortest path length of new network after wavelet decomposition is L_{new}, the average shortest path length of original network L should meet equation (6).

$$L \leq L_{new} + 2\max_{i,j} L_{ij} \tag{6}$$

3.2.3. Clustering Property Analysis

The clustering property of original network can be obtained by the lowest frequency sub-band LL_1 matrix. The value of $\tilde{a}_{ii}^r$ is much greater than that of $\tilde{a}_{ij}^r$ for all the i and j, which illustrates that the number of edges in each group is quite more than that of edges among different groups, and

means high clustering network, otherwise, the network clustering is low[15].

3.3. The Determination of the Number of Network Decomposition Levels r

For the global coupling network with an average path length of 1, there exists one edge at least between the arbitrary two groups in the new network after decomposition. It is supposed to remove all the edges between different groups, and each isolated group is regarded as a sub-network. Each group is a regular grid, and the formula (7) is expressed as follows. Where k is the mean value of each node degree[16].

$$\max_{i,j} L_{ij} \leq 2^r / k \tag{7}$$

It can be derived that L_{ij} is the logarithmic growth function of the total number of network nodes N from the formula (7), namely

$$\max_{i,j} L_{ij} \leq c \ln N \tag{8}$$

Network decomposition levels r can be obtained by formulas (7) and (8), which should meet the formula (9).

$$r \leq \ln(ck) + \ln(\ln N) \tag{9}$$

4. Empirical Research on Multi-Resolution Wavelet Analysis of Chinese Aviation Network

4.1. Network Selection and Parameter Calculation of the Network

This paper chooses the Chinese aviation transport network with relatively less number of nodes as the empirical research object. The data are from the flight schedule of Ctrip website (http://flights.ctrip.com/schedule/ScheduleIndex.aspx). The number of collected navigable cities is 163, and the number of domestic routes is 2198. And 163 navigable cities are adopted as the rows and columns of the matrix, the number of air routes between two cities is taken as the weight, then forming 163×163 aviation network adjacency matrix A.

All the parameter values of Chinese aviation network are calculated by using Ucinet software and adjacency matrix A, as follows.

(1) Network connection situation: the number of nodes is 163, the number of edges in the network is 2198, and their connection probability is 0.086.

(2) The average shortest path length of network is $L = 2.16$, namely, starting from any navigable city, it only need one transfer to reach other navigable cities.

(3) The degrees of part node cities are shown in Table 1.

Tab. 1. The degree values of part node cities.

City	Beijing	Shanghai	Guangzhou	Xi'an	Shenzhen	Kunming	Chengdu	Chongqing	Hangzhou	Qingdao
Degree	118	95	96	62	66	67	68	63	41	40

(4) Cluster analysis of the network

Clustering coefficient is the average aggregation extent of network composed by the navigable cities and adjacent cities [9], which represents the breadth of air transport. In Chinese aviation network, the clustering coefficient is 0.828, which shows the stronger clustering.

### 4.2. Empirical Analysis of the Number of the Level of Network Decomposition

In the Chinese aviation network, the maximum path length between two nodes $\max L_{ij}$ equals 4, and the number of network nodes N equals 163. The formula $c \geq 0.786$ can be obtained by formula (10).

The average value of each node degree in the network is 13.779, which is calculated by the software Ucinet. The formula (9) can be used to determine the final decomposition levels r, that is $r \leq \ln(0.786*13.779) + \ln(\ln 163) = 4.010$. There should be four-level wavelet decomposition at least in Chinese aviation network.

4.3. Four-level Wavelet Decomposition of Chinese Aviation Network
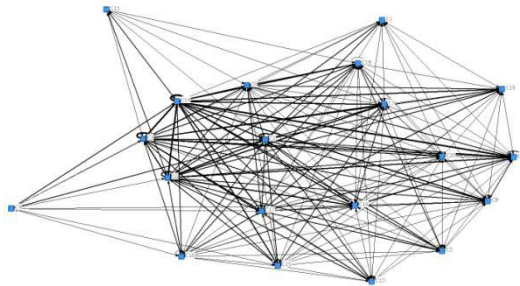
The wavelet decomposition of network is carried out by formula (6), the 163 nodes of network are divided into 16 groups, and there are 10×10 elements in each group. The four-level wavelet decomposition of network is shown in Figure 3.
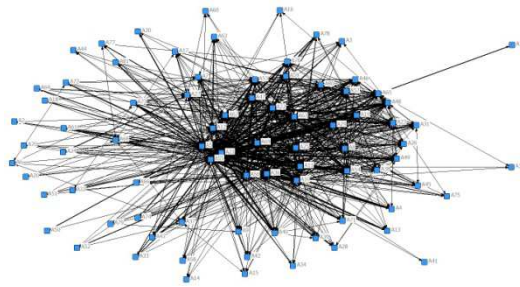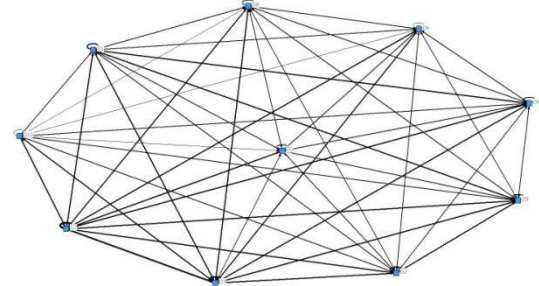

(a) The aviation network after one-level wavelet decomposition
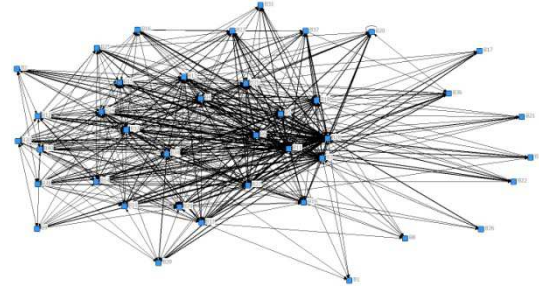
(b) The aviation network after two-level wavelet decomposition

(c) The aviation network after three-level wavelet decomposition

(d) The aviation network after four-level wavelet decomposition

Fig. 2. Four-level wavelet decomposition comparison chart.

Each element value of the lowest frequency sub-band LL_1 matrix of the network after decomposition can be obtained by the formula (7), as shown in Table 2.

Tab. 2. The adjacency matrix of Chinese aviation network after four-level wavelet decomposition.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	1.500	1.312	1.813	1.438	1.188	1.125	1.875	1.125	1.688	1.313
D2	1.375	1.750	1.500	1.313	1.375	1.063	2.250	1.125	1.625	0.813
D3	1.813	1.625	2.313	2.000	1.063	1.625	2.813	2.063	2.188	1.500
D4	1.438	1.313	2.000	0.875	0.750	0.938	2.125	1.250	1.313	0.688
D5	1.188	1.375	1.063	0.750	0.750	0.688	1.688	0.688	1.313	0.500
D6	1.125	1.063	1.625	0.938	0.688	0.500	2.188	0.750	0.750	0.500
D7	1.875	2.250	2.813	2.125	1.688	2.188	3.500	2.125	2.188	1.563
D8	1.125	1.125	2.063	1.250	0.688	0.750	2.125	0.875	1.250	0.500
D9	1.688	1.625	2.188	1.313	1.313	0.750	2.188	1.250	1.000	0.875
D10	1.313	0.813	1.500	0.688	0.500	0.500	1.563	0.500	0.938	0.188

4.4. Reduction and Contrast Analysis for the Decomposition Results of Multiresolution Wavelet

Because the wavelet transform has quite good energy tightness, the system energy mostly focuses on the low frequency sub-band, while the high frequency sub-band has a little or no energy. The information contained in the low frequency sub-band can be regarded as the rough description of the network. Therefore, the lowest frequency sub-band LL_1 matrix of aviation network after four-level wavelet decomposition contains most of the original network information, and the original network can be restored by the data in Table 2.

4.4.1. Reduction and Contrast of Network Connectivity

The nondiagonal element in Table 2 describes the connection property between ith group and jth group in original network, the greater the value of the nondiagonal element is, the larger the number of the edges between the two groups is, and the number of the original network edges equals $a_{ij}^4 \times 2^4$. For example, the corresponding element value of D2-D5 is 1.375, then the number of the edges between $2nd$ group and $5th$ group is 22 (1.375×16). Similarly, the diagonal element LL_1 describes the internal connection property of the ith group. The number of internal edges in the ith group is $a_{ii}^4 \times 2^4$, the greater the value is, the more the number of edges in the group interior is. For instance, the corresponding element value of D2-D2 is 1.750, and then the number of edges in the $2nd$ group interior is 28 (1.750×16). All the elements in Table 2 are added up, multiplied by the standardized coefficients 2^4, and the number of all the edges in the original network can be obtained. The sum of all the elements is 137.123, then is $2193.968 (137.123 \times 2^4)$ which is close to the 2198 edges in the original network. The cause of error is that there are a total of 163 nodes in the original network, cannot be evenly divisible by 2^4. When the wavelet decomposes, three nodes are ignored.

4.4.2. Reduction and Contrast of Average Shortest Path Length

The lowest frequency sub-band LL_1 is the entire connection network, and its average shortest path length is 1, which can be obtained by formulas (6) and (7).

$$L \leq L_{new} + 2 \max_{i,j} L_{ij} \leq 1 + 2 \times 2^r / k = 1 + 2 \times 16 / 13.779 = 3.32 \quad (10)$$

The average shortest path length of original network is 2.16, so the formula (10) is workable. The lowest frequency sub-band LL_1 can determine that the average shortest path value of the original network is less than 3.32.

4.4.3. Reduction and Contrast of Average Degree and Clustering Condition

the average degree and clustering situation information of the network can be restored by using the lowest frequency sub-band LL_1 of network after decomposition. The data in Table 2 can restore all the edges of the original network, a total of 2194 edges. And then the average degree of original network $k' = 2194 / 163 = 13.460$, which is close to the average value of original network degree 13.779. The difference between the minimum diagonal element 2.813 and maximum off-diagonal element 0.188 is 2.625, which is far greater than the average value 1.371 in Table 2. This indicates that the original network has high integration.

5. Conclusions

The study shows as follows.

(1) Chinese aviation network should be carried out four-level wavelet decomposition. The data in Table 2 cover the main information of Chinese aviation network characteristic.

(2) The lowest frequency sub-band LL_1 of 4 layers wavelet decomposition of Chinese aviation network includes most of the information of the original network. And the relevant information of original network can be accurately restored by adopting corresponding methods.

This paper provides an effective simplified technology for studying the quite complex network with the giant data. The subsequent study can focus on the fields such as network grouping methods, the influence of different grouping on the network information, and so on.

Acknowledgment

Reference

[1]   Liu Hongkun, Zhou Tao. Research on aviation network [J]. progress in natural science, 2008(6): 601-608.

[2]   Marr D. Vision[M]. New York: Freeman Publishers，1982.

[3]   Fan J，Wang X F. A wavelet view of small-world networks[J]. IEEE Trans. Circuits & Systems-II，2005,52(5):238-242.

[4]   Wang Xiaofan, Li Xiang, Chen Guanrong. The complex networks theory and application [M]. Beijing: Tsinghua University press, 2006

[5]   Dai Houping. A wavelet transform optimization method of complex networks [J]. Journal of Chongqing University of Science and Technology (Natural Science Edition). 2008(6):84-86.

[6]   Zhong Zhili. The study of fabric pilling objective evaluation Based on wavelet analysis [D]. Tianjin: Tianjin University of Technology, 2006

[7] Zhao Hailong, Mu Zhichun, , Ding Wenkui, Zhang Xia. Ear Recognition Based on Wavelet Transform and Block DCT[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2009 (3):243 -247.

[8] Liu Hongkun. The structure of Chinese aviation network and analysis of its influencing factor [D]. Chengdu:Southwest Jiao Tong University, 2007

[9] Chunhua Gao. Airline Integrated Planning and Operations [D].Georgia Institute of Technology，2007.

[10] Ye Wu, Ping Li, Maoyin Chen. Response of Scale-free Networks with Community Structure to External Stimuli[J].Physics A, 2009, 388 (14):2987-2994.

[11] Albert R, Barabásia L. Statistic mechanics of complex networks [J].Review of Modern Physics, 2002 (74):47-97.

[12] Kurant M, Thiran P. Extraction and analysis of traffic and topologies of transportation networks. Phys Rev E. 2006, 74:036114.

[13] Bagler G. Analysis of the airport network of India as a complex weighted network[J]. Physica A, 2008, 387: 2972-2980.

[14] Liu HK, Zhou T. Topological properties of Chinese city airline network. Dynamics of Continuous, Discrete and Impulsive Systems B, 2007, 14: 135-138.

[15] Guimerà R, Mossa S, Turtschi A. The world-wide air transportation network: Anomalous centrality, community structure, and cities'global roles. Proc Natl Acad Sci USA, 2005, 102(22): 7794-7799.

[16] Yao Hongguang, Zhu Liping. Research on Robustness of China's Aviation Network Based on Simulation Analysis[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering),2012(1):42-46.